

Konsistent vernetzte Wörterbucheinträge

PETRA STORJOHANN und PETER MEYER

Abstract

*In this paper, we address issues of inconsistencies of dictionary information and how different corpus methods and computer tools can assist in providing systematic cross-referencing. The question is raised how hyperlinking in an electronic reference work can be approached systematically in order to warrant consistent symmetrical links between synonyms or antonyms. Firstly, it is argued that working with a comprehensive corpus does not account for consistent cross-referencing. It is shown that a top-down corpus-driven linguistic analysis also does not guarantee the lexicographic documentation of binary lexico-semantic relations covered by corpus data, as proposed by Paradis/Willners (2006a, b). Secondly, with the help of dictionary examples taken from *lexiko* (an online dictionary of contemporary German) we demonstrate how a combination of both corpus-driven and corpus-based procedures enables lexicographers to systematically exploit corpus material in more depth than by using only one of these methods. It is also discussed where and why lexicographers are still prone to inconsistencies in the editing processes, irrespective of their underlying corpus methodologies. Finally, we introduce a cross-reference management tool that has been developed for *lexiko* and we explain its technological prerequisites and implications. This software supports lexicographers in detecting existing and missing references from and to a specific headword. It also offers options to automatically and comfortably correct discrepancies. Overall, we suggest a method that includes linguistic competence, complementary corpus approaches and additional software in order to ensure that links or references between synonymic and antonymic pairings are given in both directions.*

1. Einleitung

Es ist prinzipiell das Bestreben eines Lexikografen/einer Lexikografin, Wörterbucheinträge in sich konsistent anzulegen. Besonders bei einem Wörterbuch wie *lexiko*¹, das umfangreich inhaltlich zusammenhängende Informationen semantisch-pragmatischer oder diskursiver Art enthält, von Kollokationen über Synonyme und Gegensatzwörter bis hin zu regelhaften komplexeren Konstruktionen, sollten die einzelnen Angaben inhaltlich gut und nachvollziehbar miteinander verknüpft sein (vgl. dazu Klosa 2011c). Dieser inhaltlichen Konsistenz und Geschlossenheit innerhalb eines Wörterbuchartikels sollte aber ebenso eine konsequente Artikularbeit auf der Stichwortebene² gegenüberstehen, insbesondere hinsichtlich der unterschiedlichen Weise, wie einzelne Wörterbuchangaben bzw. Einträge zueinander in Beziehung stehen, z. B. durch Wortbildung oder als Teile eines konzeptuell-semantischen Feldes.

In diesem Beitrag wird die Frage beantwortet, wie in einem allgemeinen einsprachigen elektronischen Wörterbuch konsistent miteinander vernetzte Wörterbucheinträge bereits während der lexikografischen Arbeit entstehen können, welchen Anteil das systematische Arbeiten mit einem umfangreichen Korpus dabei hat und welche weiteren computerunterstützten Möglichkeiten dafür unerlässlich sind. Als ein im Aufbau befindliches allgemeinsprachliches Internetwörterbuch, dessen Artikularbeit auf der Auswertung eines umfangreichen Korpus fußt, eignet sich *lexiko* in besonderem Maße zur Klärung zentraler praktischer Fragestellungen, z. B., wie gewährleistet werden kann, dass zwischen Synonymen oder Gegensatzwörtern konsequent bidirektionale Verweise dokumentiert werden. Fragen lexikalischer Vernetzung können sich aber ebenso auf andere Wörterbuchangaben beziehen und sowohl allgemeine Bedeutungswörterbücher, Lernerwörterbücher als auch beispielsweise Synonymika betreffen. Die im Folgenden skizzierten Lösungsansätze sind analog auf andere Nachschlagewerke übertragbar, die systematisches Verweisen zwischen Stichwörtern oder zwischen bestimmten Angaben innerhalb der Wörterbuchartikel garantieren möchten.

2. Vernetzungen zwischen Stichwörtern in einem Wörterbuch

Das Wörterbuch *lexiko* ist „ein im Aufbau befindliches Online-Wörterbuch zur deutschen Gegenwartssprache, das auf der Basis eines Korpus von Grund auf neu erarbeitet wird“ (Klosa 2011a: 7). Es enthält u. a., im Unterschied zu den meisten Synonymiken oder Gegensatzwörterbüchern, Angaben zu sinn- und sachverwandten Ausdrücken, die zahlreiche verschiedene paradigmatische Beziehungsarten einschließen. In der Wörterbuchrubrik „Sinnverwandte Wörter“ werden sämtliche in paradigma-

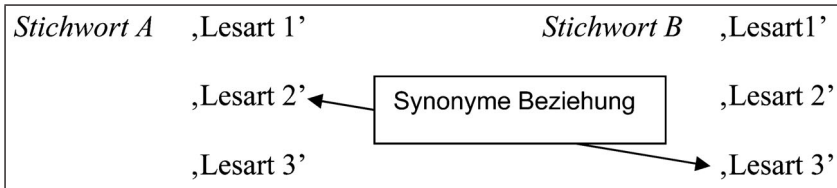


Abbildung 1. Paradigmatische Vernetzung auf Lesartenebene.

tischer Relation stehenden Ausdrücke eines Stichwortes in einem bestimmten Kontext, sprich in einer Lesart, dokumentiert, mit einem Beleg illustriert und ggf. bei bestimmten Verwendungsrestriktionen oder Kontextbeobachtungen zusätzlich kommentiert (siehe Storjohann 2005a). Die aufgenommenen Ausdrücke sind, da *lexiko* ein elektronisches Wörterbuch ist, als Link realisiert, um eine schnelle Navigation zu diesen Stichwörtern zu ermöglichen und damit auch den medialen Möglichkeiten des Internets gerecht zu werden. Das heißt, man kann problemlos von einem Stichwort in einer bestimmten Verwendung zu einem seiner Synonyme, Gegensatzwörter oder Hyperonyme/Hyponyme etc. gelangen. Die Verlinkungen sind nicht nur von Stichwort zu Stichwort angelegt, sondern auf Lesartenebene, um spezifische Beziehungen rein kontextuell zu erfassen. Damit wird gezeigt, in welchen konkreten Verwendungssituationen beispielsweise eine bedeutungsähnliche Beziehung eingegangen wird (vgl. Abbildung 1).

In einer kleinen Studie zeigt Müller-Spitzer (2010) jedoch, wie häufig lexikografische Inkonsistenzen bezüglich solch symmetrischer Beziehungen der Synonymie in gängigen Synonymwörterbüchern wie Wahrig (2006) und Duden (2007) auftreten. Anhand beispielhafter Synonymeinträge illustriert Müller-Spitzer (2010), wie typisch fehlende Rückverweise von Stichwort *A* zu Stichwort *B* oder umgekehrt sind, die vermutlich von Lexikograf(inn)en übersehen wurden. Konsistent miteinander vernetzte Synonyme in einem Wörterbuch machen aber Nachschlagehandlungen nachvollziehbarer und liegen daher im Interesse der Nutzer(innen) (vgl. Müller-Spitzer 2010: 148). Liegen diese Inkonsistenzen an den lexikografischen Auswahlkriterien für sinnverwandte Wörter oder an den genutzten Arbeitsmethoden, wie etwa bestimmten Verfahren der Korpusanalyse?

3. Ein Korpus als Konsistenzgarantie?

Insbesondere bei korpusbasierten Wörterbüchern wurde bereits die Frage formuliert, ob man bei der Aufnahme von Synonymen oder Antonymen etc. bestimmte Verfahren der Ermittlung zugrunde legen kann,

und ob die Aufnahme von sinn- und sachverwandten Ausdrücken in einem Wörterbuch bestimmten methodologischen Prinzipien folgt (Paradis/Willners 2007), mit anderen Worten, ob das Vorhandensein oder das Fehlen eines bestimmten, eventuell erwartbaren Eintrages Kriterien unterliegt, die durch die Arbeit mit Korpusdaten nachvollziehbarer sind, weil Lexikograf(inn)en stärker statistische Auswertungen sprachlicher Informationen vornehmen können. Dass die Gewinnung sinn- und sachverwandter Ausdrücke mit einem Korpus systematischer erfolgen kann und dabei verschiedene Verfahren eingesetzt werden können, wird u. a. in Storjohann (2005a, 2006) für das korpusgestützt erarbeitete Wörterbuch *lexiko* beschrieben. Dabei wird stets betont, dass natürlich die Signifikanz eine erhebliche Rolle bei der Entscheidung spielt, ob sprachlicher Usus vorliegt und daraufhin z. B. ein Synonym lexikografisch dokumentiert wird. Dennoch ist die Annahme falsch, dass allein durch die Arbeit mit einem Korpus und durch den Einsatz verschiedener Korpusmethoden die vernetzten Wortschatzstrukturen in einem Wörterbuch konsistent dargestellt werden könnten.

3.1 Fehlende Angaben im COBUILD

Paradis/Willners (2006a/b, 2007) decken im Collins COBUILD Advanced Learner's English Dictionary (2003) große Diskrepanzen bezüglich der Bidirektionalität zwischen Antonymen auf. Ihre Studie ergibt, dass nur 37 % aller Antonyme konsequent symmetrisch in beide Richtungen angegeben wurden. Besonders bei prototypischen adjektivischen Gegensatzpaaren, bei denen eine gegenseitige Vernetzung erwartet werden kann, werden die Mängel sichtbar.

Examples of pairs that occur in both directions are *dead↔alive*, *bad↔good*, *broad↔narrow* [...]. But quite unexpectedly, we found *big→small* but not *small→big* and *little→big* but not *big→little* and we found only *cheap→expensive*, *cordial→hostile*, *cruel→kind*, *difficult→easy*, *dry→sweet*, *dusk→dawn*, *old→new*, *hate→love*, *nasty→nice*, *dull→interesting*, *dull→sunny*, *dull→sharp*, *false→true*, *full→empty*, *odd→even*, *sad→happy*, *safe→dangerous*, *short→tall*, *profound→shallow*, *stale→fresh*, *dog→bitch* in that order. In our opinion as non-native speakers of English, all these pairs deserve to be reversed, e. g. *big* should be listed as an antonym of *small* and *cheap* of *expensive*. Again, it is not clear to us why these pairings should be helpful for the learner in the above directions only.

(2006a: 217)

Die Untersuchung zeigt, dass insgesamt die Aufnahmekriterien von Antonympaaren nicht transparent sind. Paradis/Willners (2006a) schlagen

vor, aufgrund des kontextuellen Verhaltens von Antonymen stärker korpusgesteuert statt korpusbasiert vorzugehen³, also die Ermittlung von Kookkurrenzen zum lexikografischen Prinzip zu erklären. Bei der korpusbasierten Methode nutzen Lexikograf(inn)en das Korpus überwiegend als Prüfinstrumentarium, um lexikografische Annahmen zu bestätigen oder zu verwerfen. In solchen Fällen werden z. B. neben der eigenen introspektiven Kompetenz Wörterbuchangaben aus bereits bestehenden Nachschlagewerken im Korpus überprüft. Bei der korpusgesteuerten Methode tritt der Lexikograf/die Lexikografin jedoch ohne Vorannahmen an Korpusdaten und wertet diese zunächst mithilfe statistischer Verfahren aus (z. B. durch eine Kollokationsanalyse). Paradis/Willners (2006a) vermuten, dass die gewählten lexikografischen Verfahren für den Grad des systematischen Verweizens zwischen binären Strukturen von Antonympaaren mit gleicher Inhaltsstruktur ausschlaggebend sind.

3.2 Fehlende Angaben in *lexiko*

Für ein systematisches Vorgehen zur Erfassung relationaler Strukturen sind die eingesetzten Korpusmethoden wichtig, auch wenn diese allein kein Garant für konsistente lexikografische Vernetzungen sind. Für die Ermittlung sinn- und sachverwandter Ausdrücke setzt *lexiko* eine Kombination aus verschiedenen komplementären Verfahren ein⁴, bei der nach mehrjähriger Erfahrung in der Praxis die korpusgesteuerte Methode die ergebnisreichste Art der Gewinnung von solchen paradigmatischen Beziehungen ist, die innerhalb eines Satzes realisiert werden. Dies betrifft vor allem Beziehungen des Gegensatzes und der Über- bzw. Unterordnung. Dieses Verfahren beruht auf der computergestützten statistischen Analyse von Kollokationen und dem Vergleichen von Kollokationsprofilen verschiedener Ausdrücke, die semantische Ähnlichkeiten aufweisen. Ermöglicht nun dieses Vorgehen, konsequent alle bidirektionalen Beziehungen in *lexiko* zu erfassen? Man kann zwar solche Beziehungen mit der korpusgesteuerten Methode besser und systematischer erfassen, aber das bedeutet nicht, dass diese ebenso konsequent bidirektional lexikografisch dokumentiert werden. Eine Auswertung des Stichwortbestandes von ca. 1.400 Stichwörtern ergibt, dass diese insgesamt über ca. 26.000 Relationspartner (das sind die Stichwörter, die eine Sinn- bzw. Sachverwandtschaft mit anderen Ausdrücken eingehen) verfügen. Von diesen Beziehungen gab es 2.658 korrekte bidirektionale Vernetzungen auf Lesartenebene, was also 1.279 Wortpaare sind. 1.700 Vernetzungen sind dagegen nicht korrekt bidirektional vernetzt⁵. Fehlerhafte Vernetzungen haben aber unterschiedliche Gründe in *lexiko*, die unter anderem auch daran liegen, dass *lexiko* kein abgeschlossenes Wörterbuch ist, sondern sich im Aufbau befindet und nicht alle Stichwörter, die ent-

halten sind, bereits vollständig erarbeitet wurden. Dazu ein Beispiel. Der Artikel *uneingeschränkt* etwa verfügt in der Lesart ‚ungehindert‘ über die Synonyme *frei*, *grenzenlos*, *unbegrenzt* und *ungehindert* (vgl. Abbildung 2).

Stichwort + Lesart		Stichwort + Lesart
<i>uneingeschränkt</i> ‚ungehindert‘:	↔	<i>frei</i> ‚unbehindert‘
	→	<i>grenzenlos</i> ‚ohne Barriere‘
	↔	<i>unbegrenzt</i>
		‚ohne Einschränkung‘
	→	<i>ungehindert</i> ‚0‘

Abbildung 2. Synonyme für *uneingeschränkt* und seine Vernetzungen.

Uneingeschränkt ist korrekt bidirektional mit *frei* und *unbegrenzt* auf Lesartenebene vernetzt. Für den Ausdruck *ungehindert* liegt keine vollständige Vernetzung vor, weil dieses Stichwort zum Zeitpunkt der Erarbeitung des Artikels *uneingeschränkt* noch nicht redaktionell bearbeitet war, also nicht über ein ausgearbeitetes Lesartenspektrum verfügte. Bei dieser Angabe handelt es sich weniger um einen Fehler als vielmehr um einen nicht vollendeten Arbeitsschritt, der erst zu einem späteren Zeitpunkt möglich ist, erst dann, wenn der Artikel selbst vollständig lexikografisch dokumentiert ist. Dagegen handelt es sich im Eintrag *uneingeschränkt* um einen Fehler, wenn für das Synonym *grenzenlos* die Rückverlinkung fehlt.

Eine mögliche Erklärung für einen solchen Fehler ist, dass die Wörterbuchartikel in *lexiko* derzeit nicht in semantischen Gruppen, sondern stärker nach Frequenzkriterien erarbeitet werden. Aber auch wenn die Abfolge der Erarbeitung stärker nach semantischen Gruppen erfolgen würde, würden Fehler dieser Art auftreten. Die schlichte Vielzahl der Vernetzungen im Bereich der sinn- und sachverwandten Ausdrücke führt dazu, dass man zum Zeitpunkt der Erarbeitung eines Stichwortes keinen Überblick darüber hat, welche der fertigen Stichwörter auf das zu bearbeitende Wort bereits verweisen. Die lexikografische Praxis zeigt auch, dass etwa für die Ermittlung von Synonymen stärker das korpusbasierte Verfahren zum Einsatz kommen muss, da mit der korpusgesteuerten Methode der Kollokationsanalyse Synonyme nicht gut erfasst werden⁶. Das liegt vor allem daran, dass Ähnlichkeitsbeziehungen häufig nicht innerhalb eines Satzes, sondern darüber hinaus realisiert werden, sinnvolle Kollokationsanalysen aber nur den unmittelbaren Kontext (i. d. R. einen Abstand von 5 Wörtern) betrachten. Das bedeutet, dass potentielle Synonyme gezielt im Korpus nachgeprüft werden müssen, was der korpusba-



Bei Beziehungen des Gegensatzes ist der Einsatz des korpusgesteuerten Verfahrens dagegen vorteilhaft. Bei sprachlichen Strukturen wie *nicht X, sondern Y* wird bei der computergesteuerten Untersuchung von X das Gegenwort Y als Kookkurrenz in Erscheinung treten und umgekehrt ebenfalls, vorausgesetzt, es handelt sich um eine typische Beziehung. Somit kann sowohl bei der Kollokationsanalyse von X ein Relationspartner Y erfasst werden und umgekehrt wird bei der Korpusuntersuchung von Y auch das in Beziehung stehende Stichwort X automatisch erfasst. Im Gegensatz zur korpusgesteuerten Arbeit begünstigt somit das korpusbasierte Vorgehen lexikografische Inkonsistenzen.

So sehr man sich hier auf redaktioneller Ebene bemüht, stets bestehende Verlinkungen zu prüfen oder ggf. nachzutragen, ein konsequentes systematisches Vorgehen ist auf diese Weise nicht möglich. Dazu müsste man das gesamte Beziehungsgeflecht eines Stichwortes und ebenso seiner entsprechenden Relationspartner in allen seinen Lesarten/Kontexten abrufbar haben. Einen Eindruck davon, welchen Umfang ein derartiges paradigmatisches Netz für einen hochpolysemen Ausdruck, wie etwa für das Adjektiv *frei*, haben kann, vermittelt Abbildung 3⁷. Im entsprechenden lexikografisch ausgearbeiteten *lexiko*-Wörterbucheintrag werden 15 verschiedene kontextuelle Verwendungen für das Stichwort beschrieben.

Bei dem in Abbildung 3 dargestellten Netz handelt es sich zudem nur um die ausgehenden Beziehungen des Stichwortes *frei*, nicht aber um die Beziehungen, die auf den Ausdruck *frei* verweisen.

Bei der kontinuierlichen redaktionellen lexikografischen Arbeit müsste also der Lexikograf/die Lexikografin ausgehende wie eingehende Beziehungen gezielt prüfen können, um symmetrische Beziehungen vor Augen zu haben und diese konsequent angeben zu können. Müller-Spitzer (2010) betont zu Recht:

As a result, the lexicographer simply forgot to add the necessary details on link reference, [...] a mistake that is bound to happen at some point when no computer support is available. It is simply impossible for lexicographers to bear all specific references in mind. This explains why it is essential to have computer assistance at hand. (Müller-Spitzer 2010: 150–151)

4. Der Vernetzungsmanager in *ellexiko*

Die Lösung für ein systematisches Verweisen/Vernetzen in diesem Bereich kann durch Computerunterstützung erreicht werden, also in Form eines Tools, das eingehende und ausgehende Vernetzungen für ein Stichwort und dessen entsprechende Lesarten prüft. Ein derartiges Werkzeug (Vernetzungsmanager) wurde im Rahmen des Projektes BZ*Velexiko*⁸ 2010 entwickelt und kommt jetzt bei der täglichen Artikellarbeit in *ellexiko* zum Einsatz.

Der Vernetzungsmanager zeigt im oberen Feld die Stichwörter, die auf den Artikel *frei* verweisen, zusammen mit der für die Vernetzung relevanten Lesart. Im unteren Feld werden die ausgehenden Beziehungen von *frei* in allen Lesarten angezeigt, die der Lexikograf bzw. die Lexikografin angelegt hat. Gleichzeitig wird zu jeder Beziehung ein entsprechender Vernetzungsstatus angegeben, der Auskunft darüber gibt, ob etwa eine Zuweisung der entsprechenden Lesart eines Stichwortes nicht erfolgt oder ob bei einem synonymen Ausdruck der Verweis in der Gegenrichtung fehlt etc. Mithilfe dieses Tools können direkt bei der Artikellarbeit fehlende Beziehungen nachgetragen, fehlerhafte Beziehungsarten korrigiert oder Links bezüglich ihres Verwendungskontextes, sprich ihrer Lesart, spezifiziert werden.

4.1 Technische Voraussetzungen

Die Realisierung einer Management-Software für Wörterbuchvernetzungen hängt wesentlich von den konkreten technischen Gegebenheiten der

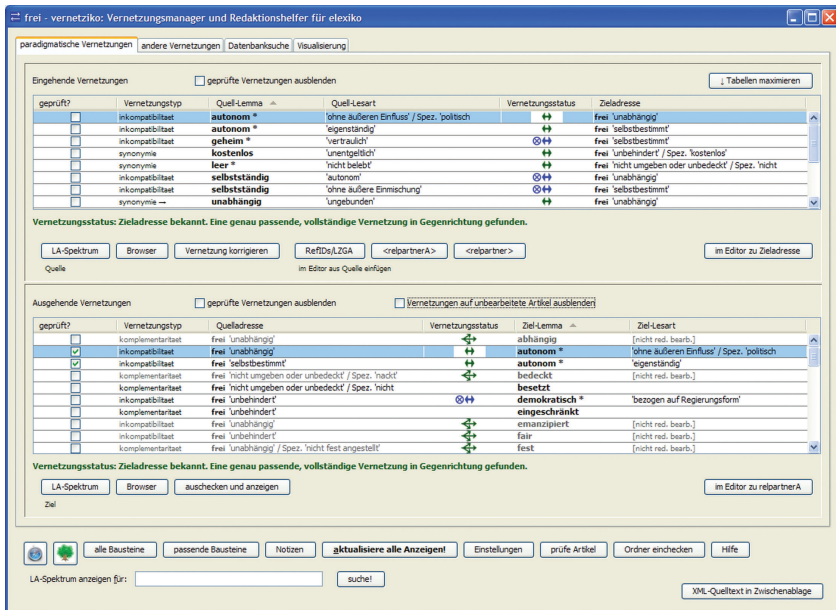


Abbildung 4. Vernetzungsmanager mit dem Artikel frei.

Redaktionsumgebung ab. Die folgenden Bemerkungen verstehen sich als kurze Skizze der Verhältnisse im *elexiko*-Projekt; für eine detailliertere Darstellung vgl. Meyer (2011). Jeder einzelne *elexiko*-Artikel ist als eigenständiges XML-Dokument in einer Datenbank abgelegt. Wenn ein Artikel zu bearbeiten ist, wird das zugehörige XML-Dokument über eine auf dem Datenbanksystem laufende Webanwendung „ausgecheckt“, d. h. als Datei lokal auf dem PC des Lexikografen bzw. der Lexikografin abgelegt und dort dann in einem XML-Editor bearbeitet⁹. Die geänderte Datei muss dann über die Webanwendung wieder in die Datenbank zurückgeschrieben („eingescheckt“) werden. Der Vernetzungsmanager ist daher auf dreierlei Weise mit der übrigen Redaktionsumgebung von *elexiko* gekoppelt (siehe Abbildung 5):

- Er muss den XML-Editor über eine vom Editor bereitzustellende Schnittstelle ‚fernsteuern‘ können, beispielsweise, um den aktuellen Bearbeitungsstand des aktuell geöffneten Dokuments auszulesen oder automatisiert Vernetzungsinformationen in das Dokument einzufügen.
- Er hat (überwiegend lesenden) Zugriff auf die Datenbank, insbesondere, um Vernetzungen auf den aktuell im XML-Editor geöffneten Artikel aus anderen *elexiko*-Artikeln zu finden.

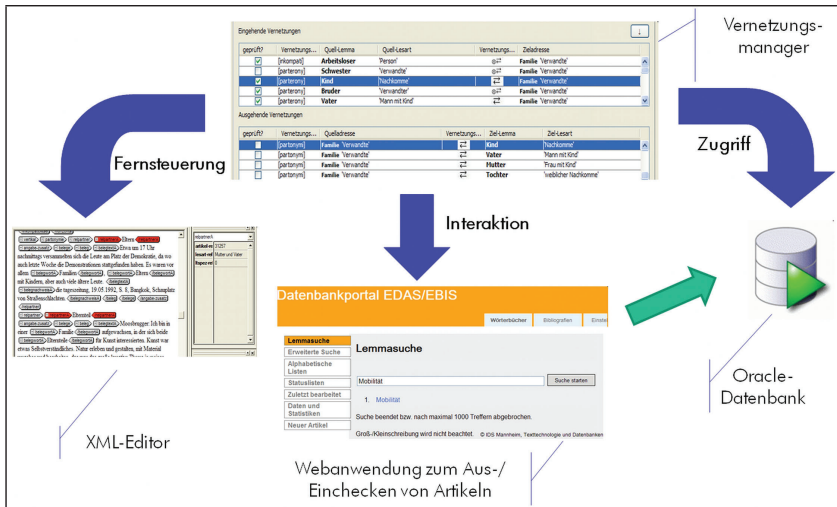


Abbildung 5. Verzahnung des Vernetzungsmanagers mit weiteren Komponenten der elexico-Redaktionsumgebung.

- c) Er kann mit der Webanwendung kommunizieren, insbesondere um automatisiert Einzelartikel oder Artikelstrecken aus- bzw. einzuchecken¹⁰.

4.2 Arbeitsweise und Funktionen des Vernetzungsmanagers

Vernetzungen werden in den *elexico*-Artikeln technisch durch XML-Elemente im verweisenden Artikel repräsentiert; das Verweisziel wird darin durch eindeutige Kennungen (Zahlen bzw. Wörter oder Wortfolgen) spezifiziert, die dem Artikel als ganzen sowie auch z. B. seinen einzelnen Lesarten über XML-Attribute zugeordnet sind (vgl. auch Müller-Spitzer 2010). Um die Bidirektionalität und Konsistenz etwa einer Antonymie-Vernetzung zu prüfen, müssen die zugehörigen XML-Elemente in den beiden Artikeln aufgesucht und die Vernetzungsinformationen miteinander verglichen werden. Dabei muss der jeweils aktuelle Bearbeitungsstand des Dokuments im XML-Editor zugrundegelegt werden.

Zu diesem Zweck liest der – in Java implementierte – Vernetzungsmanager das im XML-Editor aktuell bearbeitete XML-Dokument in Abständen von wenigen Sekunden immer wieder aus und parst es, um sämtliche ausgehenden Vernetzungen zu bestimmen. Alle Artikel, auf die verwiesen wird, müssen aus der Datenbank ausgelesen und ebenfalls analysiert werden, um die Korrektheit der Vernetzungsinformationen, insbesondere der oben erwähnten Kennungen, zu prüfen und auf diese

Weise beispielsweise Verweise zu entdecken, die „ins Leere gehen“. Artikel mit Vernetzungen auf den aktuell bearbeiteten Artikel werden ebenfalls durch eine geeignete Datenbankabfrage gefunden¹¹; alle so gefundenen Artikel müssen wiederum geparkt werden, um die konkreten eingehenden Vernetzungen zu bestimmen. Um obligatorisch bidirektionale Vernetzungstypen auf Konsistenz zu prüfen, werden abschließend die beiden Listen der ein- und ausgehenden Vernetzungen miteinander verglichen.

Die Ergebnisse der Such- und Prüfvorgänge werden in tabellarischer Form – mit Sortier- und Filtermöglichkeiten – aufbereitet. Da der Vernetzungsmanager das aktuell redigierte Dokument im XML-Editor auch verändern kann, kann der Lexikograf/die Lexikografin die jeweils gewünschten Vernetzungsinformationen automatisiert einfügen bzw. korrigieren lassen; auf diese Weise entfallen fehlerträchtige manuelle Arbeitsschritte. Auch andere Artikel, die eingehende Verweise auf das aktuell bearbeitete Dokument enthalten oder auf die vernetzt wird, können ohne Unterbrechung der Artikellarbeit mit wenigen Klicks vorübergehend im Editor geladen und korrigiert werden. Die für solche Vorgänge erforderlichen lexikografischen Entscheidungen werden interaktiv in einer Abfolge von Eingabedialogfenstern vorgenommen.

4.3 Weitere Anwendungsmöglichkeiten

Grundsätzlich können sehr verschiedene Arten von Vernetzungen zwischen Artikeln, also neben paradigmatischen Beziehungen auch z. B. aus- und eingehende Verweise auf Wortbildungsbestandteile und -produkte, softwaregestützt auf Korrektheit und Konsistenz geprüft, eingefügt oder korrigiert werden. Der Vernetzungsmanager beherrscht im derzeitigen Ausbaustadium außer den oben beschriebenen, auf Einzelartikel bezogenen Funktionen auch datenbankweite Konsistenzprüfungen und stellt diverse Hilfsmittel für den redaktionellen Alltag zur Verfügung, darunter Navigationshilfen für die Editorarbeit, frei konfigurierbare Suchen in der Wörterbuch-Datenbank sowie Visualisierung von Vernetzungsbeziehungen (vgl. Abbildung 3). Ein Ausbau um weitere Funktionalitäten ist ohne weiteres möglich.

5. Schlussbemerkungen

Bezüglich des COBUILD-Wörterbuchs lautet das Fazit von Paradis/Willners (2006a/b), dass keine transparenten Kriterien bezüglich der Auswahl der aufgenommenen Antonyme und bezüglich der Symmetrie in der Dokumentation sichtbar sind:

The principles for what antonyms are included in the dictionary are not transparent to us in spite of the fact that we have scrutinized the dictionary manually from cover to cover.

(Paradis/Willners 2006b: 104)

We fail to see a clear pattern in the choice of antonymy and the symmetry of presentation [...].

(Paradis/Willners 2006a: 216)

Für die Auswahl der sinnverwandten Wörter liegen in verschiedenen Wörterbüchern in der Regel Kriterien zugrunde. In *ellexiko* sind diese Kriterien an den sprachlichen Usus gebunden und korrelieren daher mit ermittelbarer Signifikanz im *ellexiko*-Korpus. Die Ermittlung sämtlicher sinn- und sachverwandter Ausdrücke aus diesem Korpus erfordert das Zusammenspiel korpusgesteuerter und korpusbasierter Verfahren, um optimale Erkenntnisse über paradigmatische Konstruktionen im Sprachgebrauch zu gewinnen. Aber erst mithilfe einer zusätzlichen Software kann die Dokumentation verschiedener Wortpaare und bidirektionaler Beziehungen konsistent gestaltet und nachvollzogen werden. Ein Vernetzungsmanager, wie er für *ellexiko* entwickelt worden ist, ermöglicht erst das notwendige systematische Prüfen sowie das anschließende Ergänzen und Korrigieren symmetrischer Beziehungen wie der Synonymie oder Antonymie direkt beim Verfassen eines Artikels. Er gewährleistet dann auch, dass Fehler oder inhaltliche Lücken vermieden werden können, für die man aus metalexikografischer Sicht nach vermeintlich plausiblen und transparenten Kriterien sucht.

6. Anmerkungen

1. *ellexiko* wird derzeit am Institut für Deutsche Sprache erarbeitet und ist online über das Wörterbuchportal OWID unter www.owid.de zugänglich.
2. In *ellexiko* werden Informationen zu einzelnen Wörtern gegeben. Diese Wörter, die in gedruckten Nachschlagewerken meist in alphabetischer Reihenfolge angeboten werden und auch in *ellexiko* in einer alphabetisch sortierten Liste (Stichwortliste) erscheinen, nennt man Stichwörter bzw. Lemmata. Alle Angaben des Wortartikels beziehen sich auf dieses Stichwort.
3. Die Termini *korpusbasiert* (corpus-based) und *korpusgesteuert* (corpus-driven) werden hier im engeren korpuslinguistischen Sinn verstanden und beziehen sich auf zwei speziellere Methoden der Korpusuntersuchung (vgl. hierzu Tognini-Bonelli 2001).
4. Nähere Details zu den unterschiedlichen Verfahren wurden in Storjohann (2005a, b) beschrieben.
5. Darüber hinaus gibt es weitere Vernetzungen, die bidirektional sein können, aber nicht müssen (z. B. Hyperonymiebeziehungen), und daher weder in den 2.658 noch in den 1.700 enthalten sind. Die fehlerhaften Vernetzungen in *ellexiko* sind inzwischen in einem separaten Arbeitsgang unter Einsatz der in Abschnitt 4. beschriebenen Software korrigiert.
6. Vgl. auch Storjohann (2005b).

7. Die Abbildung berücksichtigt keine Lesartendisambiguierung.
8. Zum Projekt BZVelexiko (Benutzeradaptive Zugänge und Vernetzungen in *ele-xiko*) siehe <http://www.ids-mannheim.de/lexik/BZVelexiko/>.
9. Vgl. auch Details zur Redaktionsumgebung in *ele-xiko* in Klosa (2011b: 14–16).
10. Dies ist aus technischer Sicht nicht zwingend erforderlich, da der Vernetzungsmanager das Aus- und Einchecken auch direkt mit Datenbankoperationen realisieren könnte; in der Redaktionsumgebung von *ele-xiko* ist ein solches Vorgehen aber sinnvoll, um die beim Aus- und Einchecken von Artikeln von der Webanwendung durchgeführten Prüf- und Verarbeitungsvorgänge nicht fehlerträchtig noch einmal in der Programmierung des Vernetzungsmanagers zu duplizieren.
11. Dies lässt sich technisch beispielsweise durch eine indexgestützte Volltextsuche umsetzen, vgl. Müller-Spitzer/Schneider (2009).

7. Literatur

- Belica, Cyril. 1995. Statistische Kollokationsanalyse und Clustering. Korpusanalysemodul. Mannheim: Institut für Deutsche Sprache.
- Klosa, Annette. 2011a. Vorwort. In: Klosa, Annette (Hrsg.): *ele-xiko*. Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs. Tübingen: Narr, S. 7–8.
- Klosa, Annette. 2011b. Einleitung. In: Klosa, Annette (Hrsg.): *ele-xiko*. Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs. Tübingen: Narr, S. 9–26.
- Klosa, Annette (Hrsg.). 2011c. *ele-xiko*. Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs. Tübingen: Narr.
- Meyer, Peter. 2011. vernetziko: A Management Tool for the Lexicographer's Workbench. In: Conference Proceedings of eLEX2011 (Electronic Lexicography in the 21st century: New Applications for New Users), 10–12 November 2011, Bled, Slovenia, (<http://www.trojina.si/ele2011/Vsebine/proceedings/eLex2011-25.pdf>).
- Meyer, Peter/Müller-Spitzer, Carolin. 2010. Consistency of Sense Relations in a Lexicographic Context. In: Barbu Mititelu, Verginica/Pekar, Viktor/Barbu, Eduard (Hrsg.): Proceedings of the Workshop „Semantic Relations. Theory and Applications“, 18 May 2010 at the International Conference on Language Resources and Evaluation (LREC) 2010, Malta, S. 37–47, (<http://www.lrec-conf.org/proceedings/lrec2010/workshops/W9.pdf>).
- Müller-Spitzer, Carolin. 2010. The consistency of sense-related items in dictionaries: Current Status, proposals for modelling and applications in lexicographic practice. In: Storzjohann, Petra (Hrsg.): Lexical-Semantic Relations. Theoretical and Practical Perspectives. Amsterdam: John Benjamins, S. 145–162.
- Müller-Spitzer, Carolin/Schneider, Roman. 2009. Ein XML-basiertes Datenbanksystem für digitale Wörterbücher – Ein Werkstattbericht aus dem Institut für Deutsche Sprache. In: it – Information Technology 51/4, S. 197–206.
- Paradis, Carita/Willners, Caroline. 2006a. What a corpus based dictionary tells us about antonymy. In: Corino, Elisa/Marello, Carla/Onesti, Christina (Hrsg.): Proceedings of the 12th EURALEX International Congress 2006, Turin. Band 1, Turin: Edizioni dell'Orso Alessandria, S. 213–219.
- Paradis, Carita/Willners, Caroline. 2006b. Selecting antonyms for dictionary entries: methodological aspects. In: Heinat, Fredrik/Klingvall, Eva/Manninen, Satu (Hrsg.): Working Papers in English Linguistics 6. Lund: Lund University, Department of English, S. 95–106, (http://www.sol.lu.se/engelska/dokument/wp/vol06/Paradis_Willners_06.pdf).

- Paradis, Carita/Willners, Caroline. 2007. Antonyms in dictionary entries: methodological aspects. In: *Linguistica* 61/3, S. 261–277.
- Storjohann, Petra. 2005a. Paradigmatische Relationen. In: Haß, Ulrike (Hrsg.): *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz*. Berlin/New York: de Gruyter, S. 249–264.
- Storjohann, Petra. 2005b. Corpus-driven vs. corpus-based approach to the study of relational patters. In: *Proceedings of the 3rd Corpus Linguistics Conference 2005 in Birmingham*, (<http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>).
- Storjohann, Petra. 2006. New Lexicographic Approaches to the Description of Sense Relations. In: Corino, Elisa/Marello, Carla/Onesti, Christina (Hrsg): *Proceedings of the 12th EURALEX International Congress 2006*, Turin. Band 2, Turin: Edizioni dell'Orso Alessandria, S. 1201–1212.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Wörterbücher:

- Collins Cobuild. 2003. *English Advanced Learner's Dictionary*. Sinclair, John (Hrsg.), 4. Auflage. Glasgow: Harper Collins Publishers.
- Duden Synonymwörterbuch. 2007. *Das Synonymwörterbuch*, 4. Auflage. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- elexiko. 2003 ff. *elexiko*. In: *Das Institut für Deutsche Sprache* (Hrsg.): *OWID – Online-Wortschatz-Informationssystem Deutsch*. Mannheim.
Internet www.elexiko.de oder www.owid.de (Stand März 2011).
- DUDEN-GWDS. 2000. *Das große Wörterbuch der deutschen Sprache*. 10 Bände auf CD-Rom. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Wahrig Synonymwörterbuch. 2006. *Synonymwörterbuch*. 5. Auflage. München/Gütersloh: Wissen Media Verlag.

Anschrift der VerfasserInnen:

Dr. Petra Storjohann
IDS Mannheim
R5, 6–13
68161 Mannheim
E-Mail: storjohann@ids-mannheim.de

Dr. Peter Meyer
IDS Mannheim
R5, 6–13
68161 Mannheim
E-Mail: meyer@ids-mannheim.de